Characterization of Inclusions in a Nonhomogeneous GPR Problem by Artificial Neural Networks

X. Lucas Travassos[1], Jr, D. A. G. Vieira[2], Nathan Ida[3], *Fellow, IEEE*, Christian Vollaire[4], and Alain Nicolas[4]

[1]SENAI-CIMATEC, Salvador, Brazil
[2]Federal University of Minas Gerais, Belo Horizonte CEP 31270–901, Brazil,
[3]Department of Electrical and Computer Engineering, University of Akron, Akron, OH 44325 USA
[4]Ecole Centrale de Lyon (ECL), Lyon 69006, France

This paper aims at detecting and characterizing inclusions in concrete structures by inverting ground-penetrating radar (GPR) data. First, the signal is preprocessed using the principal component analysis (PCA) and then used to train an artificial neural network (ANN). The GPR data consists of 1200 time steps. Using PCA, the data can be compressed to 286 dimensions without losing any information. Moreover, with 99.99% of the original variance the data needs only 139 dimensions. This dimensional reduction makes the ANN training easier and faster. The ANN were trained to find the buried inclusions characteristics—*depth* and *radii*—considering a nonhomogenous host medium by inverting the preprocessed data. The results show that the expected maximum error was kept under 1%, which is a remarkable result, since the host medium is nonhomogenous.

Index Terms—**Artificial neural network (ANN), buried objects, ground-penetrating radar (GPR), inverse problem.**

I. INTRODUCTION

A MAJOR technical problem when working with imaging algorithms is the large variation in the inspected target signature due to environmental conditions, geometric variations, noise, and sensors' characteristics. Yet, image classification is an important pursuit in diverse technical fields including military applications, security, and safety systems, health monitoring biomedical engineering and many others. Therefore, this can be considered as a multidisciplinary problem requiring contributions from diverse technologies.

Notwithstanding the importance of the above-mentioned applications, the nondestructive testing (NDT) of concrete requires reliable measurement methods. Pulsed radars are attractive as environment measurement methods for various applications including the examples above. The waveform data is obtained by scanning an omnidirectional antenna. The use of this waveform for estimating target characteristics is known as an ill-posed inverse problems.

In the past, various imaging or inversion techniques have been developed to refocus the scattered signals back to their true spatial location.

Most of them were based on the numerical inversion of integral equations. All these techniques are characterized by a high level of complexity, accuracy, and a significant computational burden. Consequently, the imaging of typical field data may be difficult due to problems like limited coverage, noisy data, or nonlinear relations between observed quantities and physical parameters to be reconstructed.

Therefore, it has become necessary to use more efficient analysis for interpretation of raw-data. Such analysis require algorithms by which problems having complex scattering properties can be solved as accurately and as fast as possible. This specification is difficult to achieve when dealing with iteratively solved algorithms characterized by a forward solver as part of the loop,
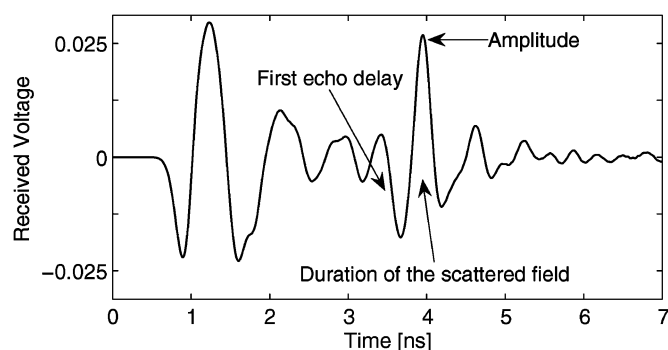
Fig. 1.   Reflected wavefield from a buried target.

which often makes the solution process computationally prohibitive for large problems. An alternative approach is the use of model-free methods based on example data. This category is represented by artificial neural networks (ANNs).

II. PROBLEM SETTINGS

The use of ANNs in the inverse scattering problem using parallel networks and networks with multiple outputs for an homogenous host medium was presented in [1] and [2]. In [2], it is shown that both configurations could deliver reasonable and very similar results using input parameters from the scattered wave defined as: 1) the peak amplitude of the reflected field; 2) the delay of the first reflected echo, calculated with respect to the time of arrival, at the receiving point, of the direct field); and 3) a measure of the duration of the scattered field (see Fig. 1). This paper considers the case when the host media is nonhomogenous and, surprisingly, a network with multiple outputs and parallel networks was not adequate to estimate the inclusion geometry.

Therefore, to solve a nonhomogenous problem the parameters presented in [2] are not sufficient. To overcome this difficulty, we implemented an algorithm to "squeeze" the scattered
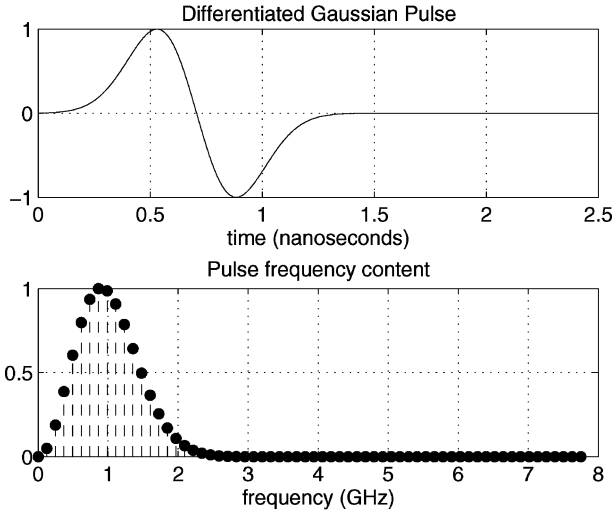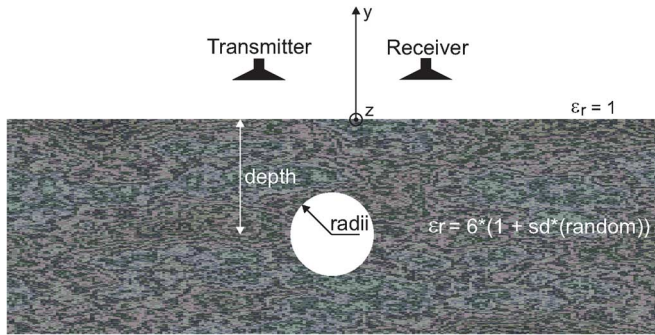
Fig. 2. Pulse description.

Fig. 3. Problem description.

wave in order to collect more information about the scatterer itself using principal component analysis (PCA).

The scattered wave is obtained using finite-difference time domain (FDTD). The problem can be summarized as the use of an incident wave and a scattered wave to characterize the scattering object. Usually, in real-world problems, the incident and scattered waves are known and it is desired to identify the scattering object.

This paper addresses a 2-D problem in which a cylinder of unknown characteristics is buried in a nonhomogeneous dielectric. The incident and scattered waves are simulated using FDTD to train the ANN. The dielectric medium uses the electrical characteristics of concrete [3] with a mean relative electrical permittivity value of 6 and standard deviation 0.15, i.e, a nonhomogenous medium. The investigation domain is illuminated by a differentiated Gaussian pulse (Fig. 2) with a center frequency of $f = 900$ MHz and bandwidth between 0.3 and 2 GHz with antennas being simulated as dipoles (Fig. 3). In order to control the numerical dispersion and provide a good discretization for the inclusions, the spatial steps were chosen as $\Delta x = \Delta y = 6$ mm. The aim of this problem is, given an incident wave, W_i, and scattered wave, W_s, to determine the *radii* and *depth* of the inclusion.

To train the ANN a set of different inclusion examples, say $S(.)$, with different *radii*, *depth*, ϵ_r, and σ were generated.

The ANN has been trained with a set of different inclusions examples, constructed by varying the *radii* in the range [0.02

$\div$ 0.1] m according to the rule $radius = 0.02 + i \times 0.001$, $i = 0, .., 80$, with ϵ_r in the range [1 $\div$ 10], according to the rule $\epsilon_r = 1 + i \times 1$, $i = 0, \ldots, 9$, σ in the range [0 $\div$ 4000] S/m according to the rule $\sigma = 0 + i \times 500$, $i = 0, \ldots, 8$ and *depth* in the range [0.05 $\div$ 0.25] m, according to the rule $depth = 0.05 + i \times 0.025$, $i = 0, \ldots, 9$ making a total of 1640 examples.

The W_s is the only information available in the real cases, therefore, it has to be used to characterize the inclusion. In this paper, 1200 time steps were considered; thus, a problem in $\mathbb{R}^{1200}$ must be solved. The ANNs suffer of a phenomenon called the curse of dimensionality, i.e., the learning process becomes slower and less effective. In the next section, the curse of dimensionality will be addressed followed by the principal component analysis (PCA). The latter is an effective way to reduce the dimensionality of this problem.

III. CURSE OF DIMENSIONALITY

One phenomenon that takes place in high-dimensional data is the sparsity of the sample points [4]. Given a data set S with T data points uniformly distributed in a p-dimensional unit sphere centered at the origin, the median distance given from the origin to the closest sample is given by

$$d(p, T) = \left(1 - \frac{1^{1/T}}{2}\right)^{1/p}. \quad (1)$$

In a sample size of 1640 and 1200 dimensions, $d(1200, 1640) = 0.9935$, which means that the samples are closer to the boundary of the space than to any other data point. For an ANFIS topology, the number of rules R for a system with p inputs and P premises is $R = P^p$; hence, it increases exponentially with the dimension p, what makes the learning slow [5]. The following section presents the PCA, which will be used for dimensional reduction.

IV. PRINCIPAL COMPONENT ANALYSIS

In some situations, the dimension of the input vector is large, but the components of the vectors are highly correlated (redundant). PCA is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. The main advantage of PCA is that once these patterns in the data are found, the data can be compressed, i.e., by reducing the number of dimensions, without much loss of information. This technique is most used in image compression.

This technique has three effects: it orthogonalizes the components of the input vectors (so that they are uncorrelated with each other), it orders the resulting orthogonal components (principal components) so that those with the largest variation come first, and it eliminates those components that contribute the least to the variation in the data set.

The input vectors are first normalized so that they have zero mean and unity variance. For PCA to work properly, one has to subtract the mean from each of the data dimensions.

The PCA uses a linear mapping of a given set of samples $S_q = \{x_1, \ldots, x_T | x_i \in \mathbb{R}^p\}$ to construct a new data set $S_p = \{y_1, \ldots, y_T | y_i \in \mathbb{R}^q\}$, where $q \leq p$.

Another interpretation of the PCA is the construction of directions that maximize the variance. The transformation V_q generates a projection space in which the covariance matrix is diagonal. The diagonal covariance matrix implies that the variance of a variable with itself is maximized and it is minimized with respect to any other variable. Thus, the q variables with higher variance in the new space should be kept. The principal components of a set of data in $\mathbb{R}^p$ provide a sequence of best linear approximations to that data, of all ranks $q \leq p$.

The problem considered is initially in $\mathbb{R}^{1200}$ but it can be projected in a $\mathbb{R}^{286}$ without any loss of information, i.e., 100% of the data variance was kept. Considering 99.99% of the variance, the variables can be projected in $\mathbb{R}^{139}$ and in $\mathbb{R}^{51}$ when 99% of the original variance is kept. These are remarkable reductions that help in reducing the curse of dimensionality.

V. EXPECTED ERROR

The model structure problem is given by choosing among a set of possible functions $f(w, x)$, $w \in \Lambda$,[1] the possible parallel layer perceptron (PLP) in the present case, the one that optimizes a given quality criterion. Mathematically

$$w^* = \arg\min_w R\left(f(w, x)\right) \qquad (2)$$

where $R(.)$[2] is a predefined quality criterion, and w^* is the argument that minimizes (2). To apply the model structure selection, it is required to define the best approximation of the desired output d; therefore, a measure of discrepancy $L(.)$ between the desired and obtained outputs should be employed. The expected risk (error) between the desired and the approximate outputs can be expressed [6] as

$$R(w) = \int L\left(d, f(x, w)\right) dF(x, d). \qquad (3)$$

The expected risk $R(w)$ measures the expected test error for the neural network, i.e., the ANN performance. The aim of the machine learning problem is to find $f(w, x)$, $w \in \Lambda$, that minimizes the risk functional presented in (3). The integral cannot be directly evaluated since the distribution $F(x, d)$ is unknown and the only available information is the training set. The training set consists of T random vectors $z_i = (x_i, d_i)$, $i = 1, \ldots, T$, independently and identically distributed (*i.i.d*) according to some unknown but fixed probability distribution $F(z)$. The training set S can be written as

$$S = ((x_1, d_1), \ldots, (x_T, d_T)). \qquad (4)$$

The risk $R(.)$ defined in (3) can be asymptotically approximated, given some consistence conditions [6], when the number of training set samples T tends to infinity. Of course, such an infinite size set is not available. To overcome this problem, resampling techniques can be used to approximated the expected risk.

The simplest resampling technique is the hold out (HO), also called external validation or validation. It consists of removing

---

[1]In the PLP case Λ is the set of real numbers but, in general, it can be a set of scalar quatities, a set of vectors, or even a set of abstract elements [6].

[2]The risk $R(w) = \int L(d, f(x, w)) dF(x, d)$ can be written using the joint probability of x and d $p(x, d)$, $R(w) = \int L(d, f(x, w)) p(x, d) dx \, dd$.
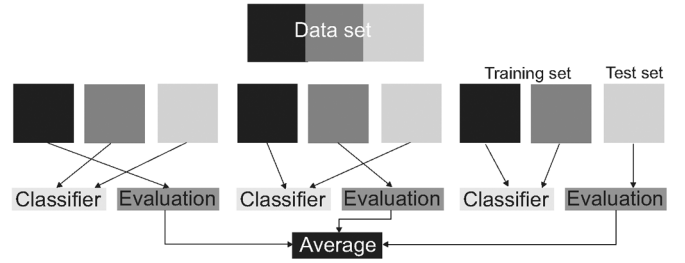

Fig. 4. Threefold cross-validation. The two sets labeled as train are used to train the neural network and the error is evaluated on the test set. After all the folders are used once as a test set, the expected risk is estimated as the average error of the sets.

TABLE I
RESULTS OF THE RELATIVE ERROR CONSIDERING THE $depth$ PREDICTION

Configuration	$mean(L)$	$max(L)$	train(s)	test(s)
$PLP(286)$	0.001%	0.01%	8.44	0.016
$PLP(139)$	0.003%	0.15%	4.48	0.01
$PLP(51)$	1%	20%	6.2	0.011

TABLE II
RESULTS OF THE RELATIVE ERROR CONSIDERING THE $radii$ PREDICTION

Configuration	$mean(L)$	$max(L)$	train(s)	test(s)
$PLP(286)$	0.022%	0.94%	7.5	0.03
$PLP(139)$	1%	20%	11.1	0.018
$PLP(51)$	1%	22%	6.1	0.013

samples from the initial learning set, and using them for validation. Many previous works used this idea of dividing the set in training and test sets. In this paper, we employed the k-fold cross-validation. For k-fold cross-validation, the training data S is divided into k sets of approximately the same size, in such a way that the learning takes place in $k - 1$ sets and the model is independently validated in the remaining set. This independence in the validation process avoids the inverse crime of using the same structure in the validation process. This is performed k times using all the k folds as validation set once. Fig. 4 shows a threefold cross-validation. The estimate of a given parameter when resampling techniques are used is the mean of the statistics evaluated on each model over the corresponding test data. The k-fold cross-validation uses the data set more effectively and is employed in this work to evaluate the expected risk.

VI. RESULTS

To evaluate the performance of the techniques studied here the following error (loss) figure is used:

$$L(d_r) = \frac{|d_t - d_r|}{d_r} \qquad (5)$$

where d is the unknown variable ($depth$ or $radii$), the subscript t indicates the real value of the variable, and subscript r indicates the value reconstructed by the neural network. This measures the percentage deviation of the reconstructed object to the real one (desired object). The expected value of the test risk, (3), given the loss functional defined in (5) was calculated using the tenfold cross-validation. Tables I and II show the mean deviation $mean(L)$, the maximum deviation $max(L)$ and the train and test times considering the dimension reduction to $q = 51$,
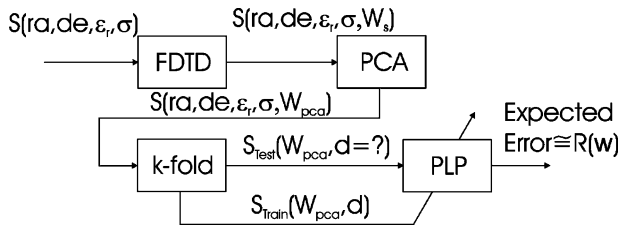
Fig. 5. Overview of the detection system employed in this paper. First, given a set S of radii (ra), depth (de), ϵ_r, and σ, a scattered wave ($W_s \in \mathbb{R}^p$) is calculated using FDTD. Afterwards, given $q < p$ a dimensional reduction is applied in W_s generating $W_{pca} \in \mathbb{R}^q$. Next, the W_{pca} is used in a k-fold system such that S_1 is used to estimate the expected error (3), and S_{k-1} are used to adapt the PLP network parameters. In this paper, a tenfold cross-validation was employed.

TABLE III
TABLE RESULTS CONSIDERING THE TWO CONFIGURATIONS
OF THE NEURAL NETWORKS STUDIED IN [2]

Configuration	$Err(depth)$	$Err(radii)$
$C1$	5.8%	12.9%
$C2$	5.7%	13.4%
$C3$	5.7%	5.8%

139 *and* 286. The system developed in this paper using a PLP network trained with the scattered wave calculated with dimensional reduction based on PCA is presented in Fig. 5.

The results of Tables I and II show that the proposed system, which combines a PCA preprocessor with a PLP is a very promising idea for assessment of inclusions in nonhomogenous concrete structures.

Retaining only 286 of the initial 1200 points, the information is still represented without any loss; thus, 100% of the data variance was kept. The reduction to 139 and 51 dimensions resulted in 99.99% and 99% of the original variance, respectively. It is clear from the results that it is harder to predict the *radii* than the *depth*. This can be physically understood by the fact that the antenna position is in front of the objects.

The nonhomogeneous problem was also solved using the approach in [2]. The two different configurations proposed to solve this inverse problem were a network with multiple outputs, and independent networks in parallel. The network with multiple outputs, called here $C1$, calculates simultaneously the *depth* and *radii* given as d_1, d_2 and d_3. The parallel networks, called here $C2$, calculate independently each output given the measured variables d_1, d_2 and d_3. In $C3$ in which the *depth* is calculated using d_1, d_2, and d_3, and the *radii* are calculated in a second step (asynchronously) using also the calculated *depth*.

As all the techniques apply a similar mechanism to reconstruct the *depth* their error are also very similar, as indicated in the second column of Table III. It can be seen in Table III that a significant error occurs when applying this technique to a nonhomogeneous problem. These results supports the initiative of using a PLP network trained with the scattered wave calculated with dimensional reduction based on PCA.

Dimensional reduction was also considered in [2] but was done empirically using only three features, the peak amplitude of the reflected field; the delay of the first reflected echo, calculated with respect to the time of arrival (at the receiving point) of the direct field and a measure of the duration of the scattered field were considered. The error average of the best configuration presented in [2] was 1.46% for the depth reconstruction that is higher than the result presented in this paper.

VII. CONCLUSION

This paper presented a novel approach of inclusion characterization in nonhomogenous concrete structures using ANN and PCA. The PCA was used to preprocess the training data aiming at dimensional reduction and, without losing any information, it was possible to represent the initial scattered wave $W_s \in \mathbb{R}^{1200}$ in a new one such that $W_{pca} \in \mathbb{R}^{286}$. This compressed wave was used to train a PLP and a low expected error was obtained with a reasonable training time.

The general system presented in Fig. 5 has proven to be effective for this type of problems and can be extended to many other GPR problems. Compared with the previous work, this paper considers the nonhomogeneity of the medium using, thus, a more realistic model. The main advantage of ANNs compared with stochastic optimization approaches is a lower computational effort for the evaluation process [7]. As it is well known, stochastic methods tend to be slow even though robust. The authors of this paper believe that the combination of both techniques must be investigated trying to extract the strengths of each one [8].

REFERENCES

[1] S. Caorsi and G. Cevini, "Neural networks trained by scattered electromagnetic data for GPR applications," in *Proc. 2nd Int. Workshop Adv. GPR*, The Netherlands, 2003, pp. 14–16.

[2] S. Caorsi and G. Cevini, "An electromagnetic approach based on neural networks for the GPR investigation of buried cylinders," *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 1, pp. 3–7, Jan. 2005.

[3] "Guidance on radar testing of concrete structures," 1997 [Online]. Available: http://www.concrete.org.uk, The Concrete Society, Tech. Rep.

[4] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning*, 1st ed. New York: Springer, Aug. 2001.

[5] J. S. R. Jang, "ANFIS: Adaptative-network-based fuzzy inference systems," *IEEE Trans. Syst., Man, Cybern.*, vol. 23, no. 3, pp. 665–685, May 1993.

[6] V. N. Vapnik, *The Nature of Statistical Learning Theory (Statistics for Engineering and Information Science)*, 2nd ed. New York: Springer, Sep. 2001.

[7] L. Travassos, D. A. G. Vieira, N. Ida, C. Vollaire, and A. Nicolas, "Inverse algorithms for the GPR assessment of concrete structures," *IEEE Trans. Magn.*, vol. 44, no. 6, pp. 994–997, Jun. 2008.

[8] D. G. Vieira, D. A.G. Vieira, W. M. Caminhas, and J. A. Vasconcelos, "A hybrid approach combining genetic algorithm and sensitivity information extracted from a parallel layer perceptron," *IEEE Trans. Magn.*, vol. 41, no. 5, pp. 1740–1743, May 2005.